Metrics for performance assessment of mixed-order Ambisonics spherical microphone arrays

Favrot, Sylvain Emmanuel; Marschall, Marton

Published in:
Proceedings of the AES 25th UK CONFERENCE

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

METRICS FOR PERFORMANCE ASSESSMENT OF MIXED-ORDER AMBISONICS SPHERICAL MICROPHONE ARRAYS

SLYVAIN FAVROT, MARTON MARSCHALL

Centre for Applied Hearing Research, Department of Electrical Engineering,
Technical University of Denmark

Mixed-order Ambisonics (MOA) combines planar (2D) higher order Ambisonics (HOA) with lower order periphonic (3D) Ambisonics. MOA encoding from spherical microphone arrays has the potential to provide versatile recordings that can be played back using 2D, 3D or mixed systems. A procedure to generate suitable layouts for a given MOA combination order is introduced consisting of rings of microphones at several elevation angles for any given MOA combination order. Robustness and directivity measures were evaluated for four MOA layouts. Results showed that MOA vertical directivity was similar to 3D HOA and that MOA horizontal directivity was in between the planar and periphonic order.

INTRODUCTION

Sound field recording techniques have seen increasing attention in the last two decades. Several applications, for example psychoacoustics and hearing instrument testing, require a realistic reproduction of these sound fields, i.e., of the spatial characteristics of the recorded scene. For these applications, high quality recordings, scalable to playback set-ups of different size, either planar (2D) or periphonic (3D), are desirable.

Higher-order Ambisonics (HOA) is a technique for either 2D or 3D systems [1] that can scale recordings from spherical microphone arrays for playback on arrays with various numbers of loudspeakers. More recently, mixed-order Ambisonics (MOA) has been introduced which combines horizontal 2D high order Ambisonics with lower order 3D Ambisonics [2]. MOA for spherical microphone arrays can theoretically improve, compared to HOA, the directivity of horizontal sources while retaining some directivity for elevated sources [3]. MOA recordings are very versatile and compatible with HOA playback. A MOA recording of combination order M_{2D}/M_{3D} could be played back (i) on either regular 3D or 2D loudspeaker arrays (using up to M_{3D} order and M_{2D} order HOA respectively) or (ii) on 3D arrays with a higher density of loudspeakers on the horizontal plane (using MOA with a combination order up to M_{2D}/M_{3D}). This requires encoded MOA signals of similar quality than corresponding M_{2D} order planar HOA and M_{3D} order periphonic HOA signals. The term "quality" here refers to (i) robustness to sensor noise and amplitude and phase mismatches and (ii) spatial resolution, i.e., directivity of the array.

This study investigates the directivity and robustness of MOA encoding from spherical microphone arrays for different combination orders, for a fixed 2D order with various 3D orders. Because of the hybrid nature of MOA, performance measures or metrics need to be evaluated separately for horizontal and vertical characteristics. First, a procedure to generate suitable microphone layouts for any given MOA combination order is introduced. Second, standard metrics are evaluated for the proposed MOA layouts in both horizontal and vertical directions. They are compared to corresponding values for 2D and 3D HOA. Finally, the effect of the regularization for MOA encoding is discussed.

BACKGROUND

First, the principle of MOA encoding using spherical arrays is briefly described here. The notations and nomenclature follow [1] and use spherical coordinates where a point in space is described by its radius r, azimuth θ ($0 \leq \theta \leq 2\pi$) and elevation δ ($-\pi/2 \leq \delta \leq \pi/2$) in relation to the origin O (the centre of the spherical array).

Pressure over a sphere

The pressure p at a point (R, θ, δ) on the surface of a solid sphere can be approximated by [1]:

$$p(kR,\theta,\delta) = \sum_{m=0}^{M} W_m(kR) \sum_{n=0}^{m} \sum_{\sigma=\pm 1} B_{mn}^{\sigma}\, Y_{mn}^{\sigma}(\theta,\delta) \quad (1)$$

with k being the wave number, the weighting factor $W_m(kR)$ for the rigid sphere as described in [1][3], B_{mn}^{σ} the Fourier-Bessel series coefficients or Ambisonics components of the sound field and $Y_{mn}^{\sigma}(\theta,\delta)$ the "real-valued" spherical harmonic functions (SHFs) [5][6] defined as:

$$Y_{mn}^{\sigma}(\theta,\delta) = \sqrt{\epsilon_n (2m+1)\frac{(m-n)!}{(m+n)!}} P_{mn}(\sin\delta) \cdot \begin{cases} \cos n\theta & \text{if } \sigma = +1 \\ \sin n\theta & \text{if } \sigma = -1 \end{cases} \quad (2)$$

with $\epsilon_n = 2, n > 0$ and $\epsilon_0 = 0$, and P_{mn} are the "Schmidt semi-normalized" associate Legendre functions of degree m and order n. For a good approximation, a large order M is used. Considering the Q pressure signals captured by microphones flush-mounted on the surface of the sphere, Eq. (2) can be written in a matrix form as:

$$\boldsymbol{s} = \boldsymbol{T}.\boldsymbol{b} \quad (3)$$

where $\boldsymbol{b}$ is the column vector of the $K = (M+1)^2$ Ambisonics components B_{mn}^{σ}, $\boldsymbol{s}$ is the Q-long column vector of the q-microphone pressure signals and $\boldsymbol{T}$ represents the transfer matrix of size $Q \times K$ written as:

$$\boldsymbol{T} = \boldsymbol{Y}.\mathrm{diag}[W_m(kR)] \quad (4)$$

with $\boldsymbol{Y}$ the SHF matrix of size $Q \times K$ evaluated at each microphone position (θ_q, δ_q).

Mixed-order Ambisonics

The MOA scheme [2] relies on a selection of SHFs. The MOA harmonic functions for a combination order M_{2D}/M_{3D} consist of all SHFs up to order M_{3D} and horizontal functions ($n = m$) of the SHFs from order $m = M_{3D} + 1$ to M_{2D}. The number of MOA harmonics K is then:

$$K = (M_{3D}+1)^2 + 2(M_{2D} - M_{3D}) \quad (5)$$

The matrix of MOA SHFs will be noted $\widetilde{\boldsymbol{Y}}$.

Encoding

MOA components are encoded from the Q microphone signals using the Ambisonic method [3]. The frequency-dependent array encoding $K \times Q$ matrix $\boldsymbol{E}(f)$ derives the coefficients $\boldsymbol{b}$ from the sampled pressures $\boldsymbol{p}$ as:

$$\boldsymbol{b}(f) = \boldsymbol{E}(f)\boldsymbol{p}(f) \quad (6)$$

and is obtained by inverting Eq. (3). Using the regularized filtering approach described e.g. [1], the encoding matrix $\boldsymbol{E}$ is approximated by:

$$\boldsymbol{E}(f) \approx \mathrm{diag}\left[\frac{W_m(kR)^*}{|W_m(kR)|^2 + \lambda^2}\right] \widetilde{\boldsymbol{Y}}^{\dagger} \quad (7)$$

where $\widetilde{\boldsymbol{Y}}^{\dagger}$ is the pseudo-inverse of $\widetilde{\boldsymbol{Y}}$ and λ the regularization parameter. The regularization prevents the classical problem of excessive amplification of high orders at low frequencies [1], which in practice would lead to the introduction of high noise level at low frequencies.

METHODS

In order to investigate the performance of MOA spherical arrays for different mixed-order combinations, suitable sensor layouts first need to be derived.

Generating ring layouts

Similarly to HOA, encoding of MOA signals relies on a least-square minimization operation (pseudo-inverse in Eq. 6). Therefore, a low condition number $\kappa(\widetilde{\boldsymbol{Y}})$ of the SHF matrix is necessary for a robust encoding of MOA signals. A pre-requisite is that the number of transducers Q is higher than the number of MOA harmonics K (Eq. 5). In practice, a higher number of sensors is needed to obtain good robustness (especially if the layout is irregular). In addition, the SHFs evaluated at sensor positions should form an orthonormal basis. For MOA, one straightforward way to achieve the orthonormality of the horizontal harmonics is to sample the horizontal ring (equator) with equiangular spacing.

The following procedure describes the generation of example layouts for a given M_{2D} and M_{3D} order. Layouts consist of rings of N_q transducers sampled at equiangular azimuth and with elevation δ_i. First, each layout includes a horizontal ($\delta = 0$) ring of $N_q = 2M_{2D} + 1$ transducers. This number corresponds to the number of horizontal SHFs up to M_{2D}. Second, the number of transducers on all elevated rings should be at least equal to N_v (to fulfill $Q > K$):

$$N_v = \begin{cases} K - (2M_{2D}+1) + 2, & M_{3D} \text{ even} \\ K - (2M_{2D}+1), & M_{3D} \text{ odd} \end{cases} \quad (8)$$

The number of elevated ($\delta \neq 0$) rings N_r is equal to M_{3D} for even 3D order and to $M_{3D} + 1$ for odd 3D order. Table 1 describes the elevation angle δ_i of ring i consisting of N_q transducers. The $\lceil . \rceil$ brackets represent the ceiling function and $\epsilon = 1$ for even M_{3D} and $\epsilon = 0$ for odd M_{3D}.

i	δ_i	N_q
$1 : N_r/2$	$\pm\left(\frac{\pi}{2}\frac{i}{N_r/2+\epsilon}\right)$	$\left\lceil\left(1 - \frac{2\delta_i}{\pi}\right)N_v + 1\right\rceil$
0	0	$2M_{2D}+1$

Table 1 – Parameters for the generation of MOA layouts

The obtained layouts are symmetric with respect to the horizontal plane. In order to verify that this procedure

provides suitable example layouts, $\kappa(\widetilde{\boldsymbol{Y}})$ was evaluated for $M_{2D} = 7$ and the considered M_{3D} orders up to M_{2D} and listed in Table 2.

M_{3D}	1	2	3	4	5	6	7
Q	17	21	25	41	51	81	103
K	16	19	24	31	40	51	64
$\kappa(\widetilde{\boldsymbol{Y}})$	1.94	2.15	1.99	1.63	1.54	1.51	1.72

Table 2 – Condition number of matrix $\widetilde{\mathbf{Y}}$

The layout obtained for a combination order 7/5 is shown in Fig. 1 for a 5 cm-radius sphere. Sensors are indicated with red circles.

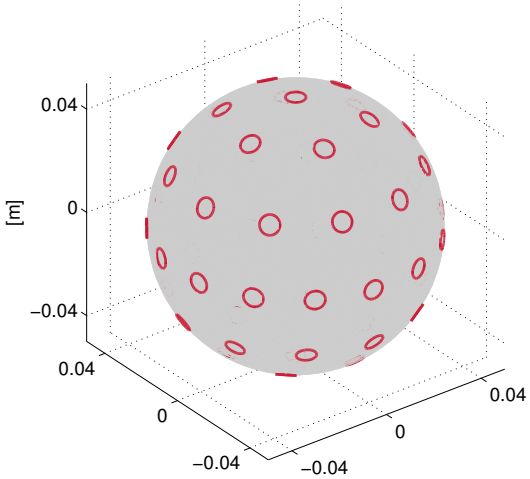

Figure 1 - Layout obtained for a combination order of 7/5.

Metrics

Since MOA does not inherently handle the horizontal and vertical directions in a similar way, array performance metrics should be evaluated for both horizontal and elevated sources and their reference values must be carefully chosen.

Using spherical microphone arrays for sound field reproduction is very similar to its use for beamforming. Established performance measures for spherical array beamforming [7] are therefore relevant for the present study and are described in the following section. For a look direction (θ_0, δ_0), the output of the beamformer can be written, for a regular beam pattern [7], as:

$$y(\theta_0, \delta_0, f) = \boldsymbol{y}_0 \boldsymbol{b}(\mathrm{f}) \tag{9}$$

With $\boldsymbol{y}_0$ being the K-long row vector of SHFs evaluated at (θ_0, δ_0) and b the obtained Ambisonics signals after array processing. Fig. 2 shows beam patterns, i.e., beamformer output plotted against direction of incoming plane wave direction, for a frontal look direction (0,0) seen from above (horizontal plane, left column) and seen from the side (vertical plane, right column) for $M_{2D} = 7$ and $M_{3D} = 1$, 3, 5 and 7 (rows). Red and blue colors indicate positive and negative values respectively. These beam patterns were obtained after simulating a large amount of incoming plane waves (3500) for a frequency $f = 5$ kHz and a radius $R = 5$ cm. Four established measures for beamformers, namely white noise gain (WNG), directivity index (DI), beam width and side lobe level are described in the following section.

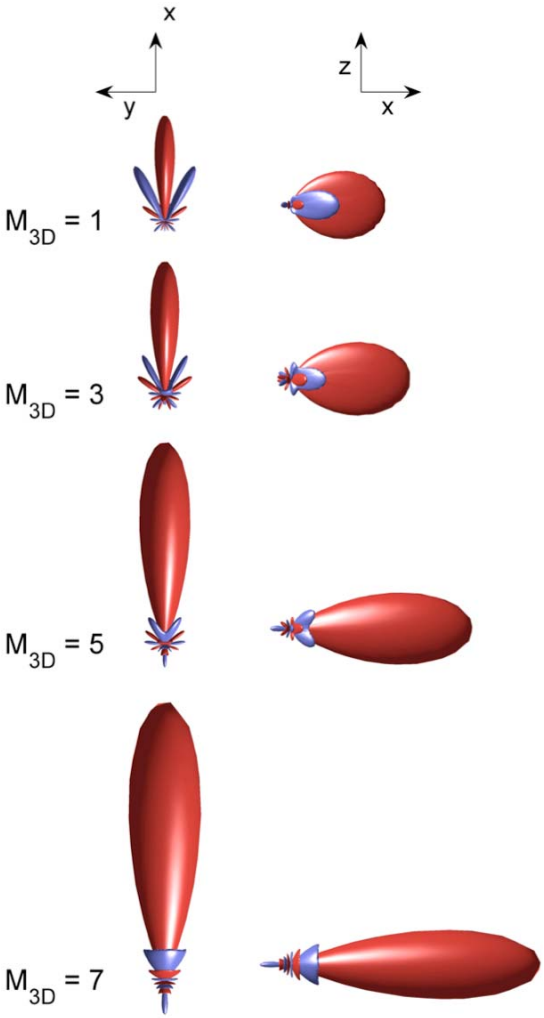

Figure 2 – Beam patterns for $M_{2D} = 7$ for a look direction $(0, 0)$ and f = 5 kHz

Another group of metrics relates to the characteristics of the reproduced sound field by loudspeakers after decoding of MOA signals. Loudspeaker array gains correspond to regular beam pattern outputs for the direction of each loudspeaker and divided by the number of loudspeakers. The reconstructed sound field is influenced by the loudspeaker layout. Therefore, in the following, a very large loudspeaker array (to avoid the influence of the playback layout) is considered to focus on the encoding effect only.

METRICS AND RESULTS

In this section, the proposed metrics for the evaluation of MOA are described and tested on 5 cm-radius hard spheres with the 4 layouts for MOA M_{2D} = 7 and M_{3D} = 1, 3, 5 and 7 (described in Table 2). The last case corresponds to 7th order 3D HOA and was used as a reference. A regularization parameter λ of -40 dB was used to investigate the effect of the regularization on MOA performances.

White noise gain

The white noise gain (WNG) is a common measure to estimate beamformer array robustness against microphone self noise, amplitude and phase variation as well as position errors [7]. WNG represents the signal energy at the output of the beamformer over the sensor self noise power as:

$$\mathrm{WNG} = 10\log_{10}\left(\frac{|\boldsymbol{y}_0\boldsymbol{b}_0|^2}{(\boldsymbol{y}_0\boldsymbol{E})^H(\boldsymbol{y}_0\boldsymbol{E})}\right) \quad (10)$$

with $\boldsymbol{y}_0$ the SHFs row vector evaluated for the look direction and $\boldsymbol{b}_0$ the MOA component encoded for the considered array with an incoming plane wave from the look direction. Fig. 3 shows WNG for a horizontal ($\delta_0 = 0$, dashed lines) and elevated ($\delta_0 = \pi/3$, solid lines) beamformer look direction for the 4 considered MOA layouts.

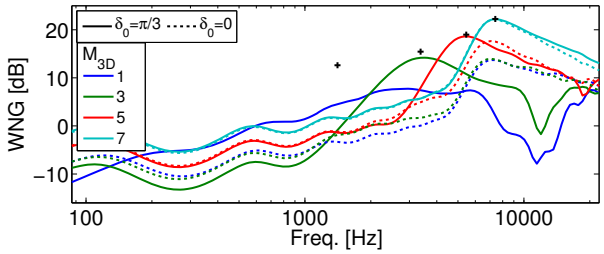

Figure 3 - White noise gain for a horizontal (dashed) and vertical (solid) beam

For comparison, a black cross marker indicates the maximum of the theoretical WNG for HOA with hard sphere arrays, which is defined as [3]:

$$\mathrm{WNG_t} = 10\log_{10}\left(\frac{Q\,K^2}{\sum_{m=0}^{M}\frac{2m+1}{|W_m(kR)|^2}}\right) \quad (11)$$

$\mathrm{WNG_t}$ presents a typical band-pass characteristic centered at a so-called optimum frequency defined as $f_{opt}(M) = cM/2\pi R$ [8] with c the sound velocity (here, $f_{opt}(7) = 7643$ Hz). For MOA arrays, WNG for the horizontal beam showed a band-pass characteristic centered at $f_{opt}(7)$ albeit with maximum values corresponding to M_{3D} HOA. For the elevated beam ($\delta_0 = \pi/3$), WNG matched the corresponding theoretical WNG for HOA of order M_{3D} except for $M_{3D} = 1$, which showed significantly lower values. This highlights that the robustness of MOA encoding matches the expected M_{3D} order HOA robustness for elevated directions and the M_{2D} order one for horizontal directions for $M_{3D} \geq 3$.

Directivity index

The directivity index indicates how directive a beamformer is, which is directly linked to its order. It is defined as the ratio of the beam output looking at the incoming plane wave relative to the average output of the same beamformer for all incoming directions [7]:

$$\mathrm{DI} = 10\log_{10}\left(\frac{|\boldsymbol{y}_0\boldsymbol{b}_0|^2}{\sum_{n=1}^{L}|\boldsymbol{y}_0\boldsymbol{b}_n|^2}\right) \quad (12)$$

where $\boldsymbol{b}_n$ is the encoded MOA signal for an incoming plane wave n out of L waves from full 3D incidence. DIs were computed after the simulation of $L = 150$ plane waves for a horizontal and elevated beam direction and are plotted in Fig. 4.

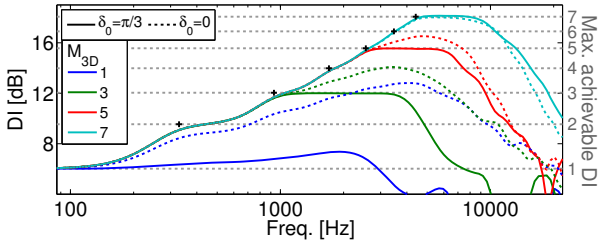

Figure 4 - Directivity index for elevated and horizontal beams

Horizontal gray dashed lines represent maximum achievable DI (DI_{max}) values for a regular beamformer of order M = 1 to 7 given as [7]:

$$\mathrm{DI}_{max}(M) = 20\log_{10}(M+1) \quad (13)$$

The frequency $f_{act}(m)$ at which the order m is "activated" by the regularization of parameter λ (Eq. 7) is defined as the frequency for which $\lambda = W_m(kR)$. Black crosses indicate points at $(f_{act}(m), \mathrm{DI}_{max}(m))$.

For the elevated beam direction, DIs increased with frequency until the maximum achievable DI for M_{3D} $\mathrm{DI}_{max}(M_{3D})$, which was reached at $f_{act}(M_{3D})$ (black crosses). This DI increase at low frequencies results from the applied regularization, which reduces the effective order of the array to avoid noise amplification for practical arrays. DI dropped above $f_{opt}(M_{3D})$ where spatial aliasing reduced directivity. It should be noted that for $M_{3D} = 1$, DI reached slightly higher values than $\mathrm{DI}_{max}(1)$. For the elevated sources, DI values did not reach $\mathrm{DI}_{max}(M_{2D})$. Instead, DI values were slightly lower than the mean of $\mathrm{DI}_{max}(M_{2D})$ and $\mathrm{DI}_{max}(M_{3D})$. Similarly, the frequency at which DI drops is roughly the mean of $f_{opt}(M_{3D})$ and $f_{opt}(M_{2D})$. MOA encoding provides similar directivity to M_{3D} order HOA for elevated directions. For horizontal directions, DI for

MOA is half way between the directivity of M_{3D} and M_{2D} order HOA.

Beam width

The main beam width of beam patterns measures the spatial resolution of the beamformer. This measure focuses only on the main beam and therefore partially describes the directivity of the beam. The 3 dB beam width is commonly defined as the angular width of the main lobe of the beam pattern at the -3 dB points relative to the maximum. This measure can conveniently be evaluated along the azimuthal and elevation direction, making it a relevant metric for MOA.

Azimuthal and elevation beam widths were computed for a frontal look direction $(0, 0)$ for the above-mentioned MOA layouts and are shown in Fig. 5.

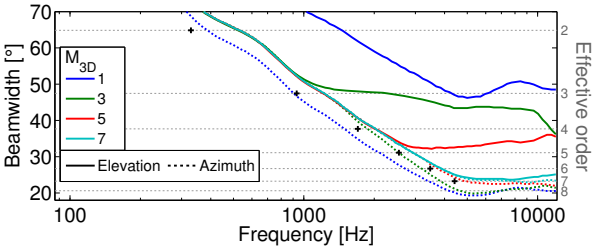

Figure 5 - Azimuthal and elevation beam width

For comparison, "effective" orders representing the beam width for ideal HOA of order 2 to 8 are plotted as dashed horizontal and black crosses were marked at $f_{act}(m)$. Beam widths are directly visible in Fig. 2 where beam patterns shown for $f = 5$ kHz. Beams get narrower as frequency increases when the regularization is effective. At $f_{act}(M_{3D})$, elevation beam widths reached the ideal width for M_{3D} except for $M_{3D} = 1$ that reached the width of an ideal 3rd order HOA beam. Surprisingly, azimuthal beam widths were slightly narrower for M_{3D} = 1 and 3 compared to higher order 5 and 7. MOA beams in horizontal look directions have an azimuthal width of M_{2D} HOA and an elevation width of M_{3D} HOA, except when for small M_{3D}.

Side lobe levels

Beam patterns usually consist of a main lobe in the look direction and of several side lobes of lower level (as seen in Fig. 2). Side lobe level (SLL) is a relevant measure to analyze alongside beam width since the spatial sensitivity of the beam (i.e., of the array) is greatly impaired by high SLLs. A local maximum searching algorithm was used to calculate SLLs for the previous simulation conditions (Fig. 6).

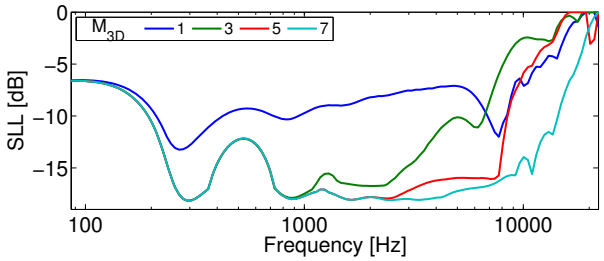

Figure 6 - Side lobe levels relative to the beam maximum

Results showed minima at $f_{act}(m)$ for $m = 2$ and 3 (below 1 kHz), which are due to the applied regularization. For $M_{3D} = 1$, SLLs present large values compared to higher 3D orders. For higher M_{3D}, SLLs increased for frequencies proportional to M_{3D} and upward, highlighting the spatial aliasing effect.

Sound field reconstruction

Sound fields are reconstructed by decoding MOA signals onto a loudspeaker array. Comparison of reconstructed versus original sound fields can provide a measure of array encoding performance when using a very large loudspeaker array. This comparison relies here on the norm of the energy vector r_E, a concept proposed by Gerzon [10]. This metric quantifies, on the playback side, the amount of energy coming from the expected direction and indicates the directivity of the reproduced sound field. For ideal 3D HOA, $r_E = M/(M+1)$ according to [6]. The transform $M_{eff} = r_E/(1 - r_E)$ can be used to relate to the effective order M_{eff}.

In the following, a virtual, 204-element regular loudspeaker array, based on a spherical t-design [9] was used to decode MOA signals encoded by the above-mentioned MOA microphone arrays. r_E values were calculated after simulation of a horizontal and $(\delta_S = 0)$ an elevated $(\delta_S = \pi/3)$ source for the four considered microphone arrays. The corresponding effective order M_{eff} is shown in Fig. 7.

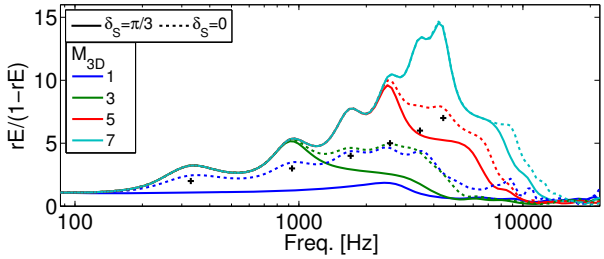

Figure 7 - Effective order transformed from the energy vector norm r_E

Black crosses indicates points at $(f_{act}(m), m)$ representing the activation of each order m by the regularization. Effective order reached higher values than M_{2D} and M_{3D} at frequencies around $f_{act}(m)$ and

showed several peaks at $f_{act}(m)$. The regularization approach applied here had a strong impact on directivity measured by r_E. The transformation to effective order is used here to ease the comparison between the different MOA combination order. It is however unclear how r_E can reach higher values than ideal HOA when regularization is applied. The directivity dropped for the elevated source for frequencies above $f_{act}(M_{3D})$. Reproduced horizontal sources showed a higher directivity for a wider bandwidth than elevated sources. It should be noted that the effective order for $M_{3D} = 1$ was similar to that of $M_{3D} = 3$.

DISCUSSION

Various metrics have been evaluated for the set of MOA layouts considered here. Separate evaluation along horizontal and vertical directions provided a tool to assess the performance of MOA in term of robustness and directivity. For the latter characteristic, the directivity index aggregated results from the beam width and side lobe levels. On the reproduction side, the norm of the energy vector r_E provided another measure of directivity.

However, the regularization affected the r_E results. Therefore, another simulation was performed without regularization to assess its effect. Since simulations only involve a target plane wave and no noise or microphone characteristic variations, obtained loudspeaker gains did not show large values. Results are shown in Fig. 8 for the same conditions of Fig. 7.

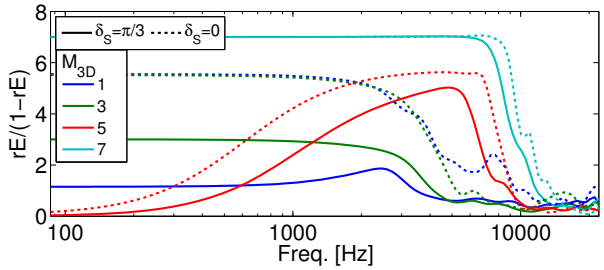

Figure 8 - Transform of the energy vector norm without regularization

Effective orders did not exceed M_{3D} values for elevated sources. For horizontal sources, the maximum effective order was 5.6 for all MOA conditions. The effective order for $M_{3D} = 5$ dropped as frequencies decreased below 5 kHz. Further investigations are needed to explain the results for this specific condition. The regularization filtering approach has an important impact on r_E at frequencies below and slightly above f_{act}. Such impact was not observed in the DI values, suggesting that DI is better suited for evaluating array directivity.

An alternative to the regularization scheme needs to be investigated for the MOA case. One approach could make use of signal-to-noise ratio measures to adapt the regularization to the recording conditions. Moreover, the impact of the present and other approaches of regularization on perception of the reproduced sound field should be investigated.

CONCLUSIONS

The present study presents metrics for assessing the performance of MOA spherical microphone arrays. A procedure was introduced to generate sensor layouts for any MOA combination order. To highlight MOA properties, encoding performance can be assessed by (i) evaluating beamformer metrics for horizontal and elevated beam separately and by (ii) evaluating the reproduced sound field for a horizontal and vertical source separately. Selected metrics were evaluated on 4 layouts. Results showed that the elevation characteristics of MOA encoding are similar to that of 3D HOA of corresponding order, whereas the horizontal characteristics are half way in between that of HOA of corresponding planar and periphonic order. Moreover, the proposed procedure provided suitable layouts for MOA. Finally, the regularization approach had a significant impact on reproduced sound fields from MOA recordings.

REFERENCES

[1] Moreau, S., Daniel, J. and Bertet, S., “3D sound field recording with higher order Ambisonics – objective measurements and validation of a 4th order spherical microphone,” presented at the AES 120th convention, Paris, France, 2006.

[2] Favrot, S., Marschall, M., Käsbach, J., Buchholz, J. and Weller, T., “Mixed-order Ambisonics recording and playback for improving horizontal directionality,” presented at the AES 131st convention, New York, USA, 2011.

[3] Marschall, M., Favrot, S., Buchholz, J., “Robustness of a mixed-order Ambisonics microphone array for sound field reproduction,” presented at the AES 132nd convention, Budapest, Hungary, 2012.

[4] Rafaely, B., “Analysis and design of spherical microphone arrays,” IEEE Transactions on Speech and Audio Processing, 13(1), pp. 135– 143, 2005.

[5] P. Morse, H. Feshbach: Methods of theoretical physics. McGraw-Hill, 1953.

[6] Daniel, J. (2000), “Representation de champs acoustiques, application a la transmission et a la reproduction de scenes sonores complexes dans un contexte multimedia,” Ph.D. thesis, 1996-2000, Universite Paris 6.

[7] Meyer, J. and Elko, G., “spherical microphone arrays for 3D sound recording,” in audio signal processing for next-generation multimedia communication systems, 2004.

[8] Park, M. and Rafaely, B., “Sound-field analysis by plane-wave decomposition using spherical microphone array,” Journal of the Acoustical Society of America, 118(5), pp. 3094–3103, 2005.
[9] Hardin, R.H. and Sloane, N.J.A., Spherical Designs, http://www2.research.att.com/~njas/sphdesigns/, last retrieved 15 Jan 2012.
[10] Gerzon, M. A., “General Metatheory of Auditory Localisation,” presented at the AES 92nd convention, Vienna, Austria, 1992.